NASA TECHNICAL MEMORANDUM

NASA TM X–62,090

NASA TM X–62,090

BLOCK UP-AND-DOWN DESIGN FOR ESTIMATING SENSORY THRESHOLDS

by Don H. Card and Michio Aoyagi

Ames Research Center
Moffett Field, Calif. 94035

October 1971

BLOCK UP-AND-DOWN DESIGN FOR ESTIMATING SENSORY THRESHOLDS

DON H. CARD AND MICHIO AOYAGI

Ames Research Center, NASA, Moffett Field, California 94035

SYMBOLS

C_n	sum of the observed stimulus levels in first n blocks (see Eqs. 2 and 3)
d	step size (distance between adjacent stimulus levels)
E	efficiency (see Eq. 9)
F(x)	probability of correct response due to sensory mechanism (see Eq. 4)
G(x)	probability of correct response due to guessing (see Eq. 4)
k	trial block length
n	number of trial blocks
P(x)	probability of correct response (see Eq. 5)
P_-	probability of decreasing the stimulus level following a block of trials
P_+	probability of increasing the stimulus level following a block of trials
P_o	probability of remaining at the same stimulus level following a block of trials
r	decrease stimulus level if number of correct responses in a block is $\geq$ r
RE	relative efficiency (see Eq. 10)
s	increase stimulus level if number of correct responses in a block is $\leq$ s
s	a general fixed stimulus level
x_i	ith stimulus level fixed prior to an experiment
y_j	stimulus level used in (j+1)st trial block
y_o	initial stimulus level
ϕ	phasing factor; distance from μ to the fixed stimulus level closest to μ
μ	true threshold value

$\hat{\mu}$ estimator of μ (Eq. 1)

π_i asymptotic frequency of stimulus level i

σ standard deviation of the underlying normal distribution $F(x)$

SUMMARY

The block up-and-down, two-alternative, forced-choice experimental design for the estimation of sensory thresholds is investigated. A mathematical model of the procedure is developed and recursion formulas are derived for bias and mean-square error of the sample average estimator. Block designs for various step sizes are compared on the basis of two measures of efficiency: (1) efficiency expressed as the reciprocal of the mean-square error of the average estimator per trial and (2) relative efficiency of the average estimator with respect to the probit estimator based on fixed stimulus levels.

INTRODUCTION

Up-and-down or "staircase" designs have been extensively analyzed and applied during the past few years to psychophysical experiments.[1-8] These experimental designs, widely used in other fields such as bioassay and explosives research,[9-13] were developed to estimate points on a response function where responses are quantal; that is, responses are categorized as occurring or not occurring following a stimulus. In the application considered in this report, this function is the probability of correct response and is assumed to be in the form of a gaussian psychometric function adjusted by a correction for guessing. The up-and-down method consists in giving the experimental subject a

series of stimuli at a fixed number of sequential trials according to the following rule: (a) following a nonresponse, increase the stimulus to the next higher stimulus level for the next trial, (b) following a response, decrease the stimulus to the next lower stimulus level. This procedure tends to concentrate testing around a fixed point on the response function whose corresponding stimulus level may be defined as a "threshold," or point of subjective equality. The block up-and-down method is an extension of the classical up-and-down method in which the decision to raise or lower the stimulus level is based on the outcome of a block of several trials rather than on the outcome of just one trial.

Recent work in this area has been concentrated on developing efficient sequential methods - experimental designs in which the number of trials is a random variable that depends on the past history of the trial sequence.[6,9,12,14,15] Nonsequential designs are sometimes easier to apply experimentally, however, and are still used in psychometric testing.[1,4,5]

Up-and-down designs, when used to determine sensory thresholds for human subjects, are usually applied within the framework of a two-alternative, forced-choice procedure; that is, the subject is required to give one of two types of response to a stimulus presented in the two categories with equal probabilities. For example, Clark and Stewart,[5] in a study of angular acceleration, required a subject in a centrifuge to respond right or left, depending on which direction he subjectively perceived that he was accelerating. The

difficulty in using an up-and-down design with the restriction of forced choice is that the probability of responding "correctly" to a stimulus varies from 0.5 to 1 as the stimulus level is increased, rather than from 0 to 1 which is assumed in other contexts, such as bioassay.[10,12,13] The probability of 0.5 at zero stimulus level arises from the assumption of random guessing under the restriction of an equally probable forced (binary) choice.

Experience with empirical response curves has shown that this restriction results in a probability-of-correct-response function that is not symmetric about the threshold; that is, the probability of increasing a step when below threshold is not equal to the probability of decreasing a step when above threshold. This lack of symmetry can also be deduced from the mathematical model to be developed below. Since classical up-and-down designs assume symmetry of the response function, we undertook to determine how the asymmetry induced by the forced-choice technique affects bias and precision of the average estimator of threshold. This method involves averaging observed stimulus levels over a series of trials[10,13] to estimate threshold, and is found quite frequently in the psychometric literature.[1,4,5,7,8]

This report summarizes an investigation of the bias and precision of the average stimulus level estimator (the dose average estimator of Ref. 13) for a block up-and-down, two-alternative, forced-choice design (BUDTIF). A mathematical model of the forced-choice procedure is developed, and well-known recursion formulas for exact bias and

mean-square error of the average estimator are modified to handle the forced-choice case. Since this was not intended to be an extensive parametric study, the main results are presented in terms of optimal block design, that is, the number of trials at a given stimulus level and the appropriate decision procedure to raise or lower the following stimulus level, which maximizes some measure of efficiency of the average estimator. The measure of efficiency chosen, and one that incorporates the influence of most of the relevant parameters, is the reciprocal of the mean-square error of the average estimator per trial.

The parameters that influence the up-and-down procedure are: (1) initial stimulus level, (2) step size (i.e., the fixed distance between stimulus levels), (3) number of trial blocks, (4) block design (described below), and (5) phasing factor. All these parameters, except the phasing factor, will be discussed. The phasing factor, which is the distance from the threshold to the stimulus level nearest the threshold, was found to have a slight but insignificant effect on mean-square error and bias for moderate trial sequence lengths. Because of this small effect on mean-square error and bias of the threshold estimator, and since the phasing factor cannot be known to the experimenter in practice, all results are presented for a phasing factor of zero.

I. MODEL OF THE EXPERIMENTAL PROCEDURE

The general psychophysical method of interest is an expansion of the BUDTIF procedure developed by Campbell[8] and the multiple up-and-down (MUD) procedure used in bioassay.[11,12] The rules for manipulating the independent variables are:

1. Choose a set of stimulus levels that are equally spaced (usually in log units of physical magnitudes).
2. Perform a sequence of trials in blocks of length k. After each block of trials is completed at a given stimulus level, select the stimulus level for the next block as follows:
   a. Increase to the next higher level following s or fewer correct responses in the present block.
   b. Decrease to the next lower level following r or more correct responses ($r > s$).
   c. Remain at the same stimulus level following a number of correct responses between s and r (not including s or r).
3. Terminate the experiment after n blocks of trials.

These parameters determine the block design (k, s, r).

This rule for changing the stimulus level after a trial block is more general than Campbell's in that he considered only the case in which the decision to raise, lower, or keep the stimulus level the same was based on whether the proportion of correct responses (x/k) was less than, greater than, or equal to the desired proportion (p) of correct responses, respectively. In confining himself to this rule, he assumed that the condition $x/k = p$ had to be a possible outcome, where x is

the number of correct responses in a block of k trials, and p is some target percentage correct which is being tracked. This condition necessarily confined him to block lengths that are multiples of 4 when $p = 0.75$, an unnecessary restriction when the above rule is used.

Various suggestions have been made concerning the best way to avoid the bias that results in the threshold estimate when the first stimulus value is far away from the true threshold. Campbell[8] suggested using only those levels, in the threshold calculation, that have been used at least twice within the trial sequence; Brownlee[13] suggested calculating threshold from those levels used after the first reversal (change of direction) of the staircase; Hsi[12] suggested beginning a sequence with single-trial blocks and then switching to k-trial blocks after the first stimulus reversal. Since all these methods are designed to place the initial trial block for the calculation of the threshold estimate in the vicinity of the threshold, a suggestion of Hsi[12] was followed and in the computer study the starting stimulus level was confined to within three standard deviations of the true (simulated) threshold.

II. STATISTICAL METHOD

Recursion Formulas for Bias and Mean Square Error

The average estimator is defined as follows: Let x_i be the ith stimulus level, fixed prior to the experiment, $i = 1, 2, \ldots,$ where $x_{i+1} > x_i$ for all i, and let y_j, $j = 0, 1, 2, \ldots,$ be

the stimulus level used in the $j+1$ trial block. If μ is the true threshold, consider estimating μ by $\hat{\mu}$, where

$$\hat{\mu} = \sum_{j=1}^{n} ky_j/nk = \sum_{j=1}^{n} y_j/n \qquad (1)$$

and k is the number of trials per block in a sequence of n blocks. As in Reference 13, the initial stimulus level y_o is not included since it was chosen by the experimenter, and the level y_n is included since it was selected on the basis of the experiment.

Recursion formulas for the exact bias and mean-square error of $\hat{\mu}$ are given by Hsi.[12] Hsi generalized to the case of $k > 1$ the formulas developed by Brownlee et al.[13] for $k = 1$. The only modification of Hsi's formulas that is required is the definition of P_+, P_o, and P_-, the probabilities of increasing the stimulus level after a block of trials, staying at the same level, and decreasing the level, respectively. These probabilities will be derived later. The recursive formulas for the bias and mean-square error, as given in Reference 12, are as follows (given that $\mu = 0$):

$$E[C_{n+1}(y)|y_o=x_i] = E[C_1(x_i)] + P_+E[C_n(x_{i+1})] + P_oE[C_n(x_i)] \qquad (2)$$

$$\begin{aligned} E[C_{n+1}(y)|y_o=x_i^2] = {} & E^2[C_1(x_i)]+P_+E[C_n^2(x_{i+1})]+P_oE[C_n^2(x_i)] \\ & +P_-E[C_n^2(x_{i-1})]+2kx_i\{E[C_{n+1}(x_i)]-E[C_1(x_i)]\} \\ & +2k\{(x_{i+1}-x_i)P_+E[C_n(x_{i+1})]-(x_i-x_{i-1})P_-E[C_n(x_{i-1})]\} \end{aligned} \qquad (3)$$

where

$$C_n(y) = \sum_{j=1}^{n} ky_j$$

A computer program was written to compute the exact bias and mean-square error using the above formulas. On a "per trial" basis, the result is

$$\text{Bias} = E[C_{n+1}(y)|y_o=x_i]/kn$$

$$\text{MSE} = E[C_{n+1}(y)|y_o=x_i]^2/k^2n^2$$

Statistical Model

The probabilities P_+, P_o, P_- may be derived as follows: In a forced-choice procedure, a correct response at a given stimulus level x is the result of a correct response due to the actual sensory mechanism under study, or is due to some random response process. If the probabilities associated with these two events are represented by $F(x)$ and $G(x)$, respectively, and if we represent the probability of a correct response by $P(x)$, then

$$P(x) = F(x) + [1 - F(x)]G(x) \qquad (4)$$

where $0.5 \leq G(x) \leq 1$ and $0 \leq F(x) \leq 1$. The psychological interpretation of this equation is: with probability $F(x)$, the subject detects the signal (perceives the stimulus); with probability $1 - F(x)$, he does not, and if he does not then he responds with a probability of $G(x)$. The inequalities imply that $0.5 \leq P(x) \leq 1$. The case considered in this report occurs when $G(x)$ is independent of x and is

completely random (i.e., guessing). In such case, $G(x) = 0.5$ for all x and $P(x) = [1 + F(x)]/2$. The term $F(x)$ is referred to in the literature as the psychometric function and is commonly assumed to be described by the normal cumulative distribution. The same assumption is made here; thus,

$$P(x) = \frac{1}{2}\left\{1 + \int_{-\infty}^{x} (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt\right\} \qquad (5)$$

Since $F(x)$ represents the actual sensory mechanism under study, by definition its mean (in this case also its median) μ is the threshold value of interest. If μ is substituted for x in Eq. 5, $P(\mu) = 0.75$. This is the reason for the 75% target percentage correct that is commonly estimated in forced-choice situations.[2]

Now, the probability of increasing, decreasing, or keeping the same stimulus level following a trial block may be represented by the binomial distribution partial sums. Thus,

P_+ = probability of increasing the stimulus level following a block of length k

= probability of s or fewer correct responses in the block

$$= \sum_{m=0}^{s} \binom{k}{m} [P(x)]^m [1 - P(x)]^{k-m}$$

P_- = probability of decreasing the stimulus level following a block of length k

= probability of r or more correct responses in the block

$$= \sum_{m=r}^{k} \binom{k}{m} [P(x)]^m [1 - P(x)]^{k-m}$$

P_o = probability of remaining at the same stimulus level following a block of length k

= probability of between s and r correct responses

$= 1 - P_+ - P_-$

Without loss of generality, the standardized form is assumed (μ and σ were taken to be 0 and 1, respectively). The step size d was taken to be constant for a given trial sequence ($d = x_i - x_{i-1}$ for all i).

Asymptotic Frequency Distribution of Stimulus Levels

The up-and-down method can be thought of and modeled as a random walk on the real line with fixed step size. This way of treating the problem is useful for deriving the asymptotic frequency distribution of stimulus levels, that is, the relative frequency with which each stimulus level is visited in an infinitely long trial sequence. Tsutakawa[11] derived this distribution for the bioassay case, and the only modification of his formulas that we require is the interpretation of P_+ and P_-, the probabilities of increasing a step and decreasing a step. As shown above, these are the probabilities of the tails of a binomial density function.

If we let p^* be the value of p that makes $P_+ = P_-$, then it can be shown that for $p^* > 1/2$:

$$P_+ > P_- \quad \text{if} \quad p < p^*$$

$$P_+ < P_- \quad \text{if} \quad p > p^*$$

indicating that the stimulus series will not drift to plus or minus infinity, and that the asymptotic distribution of stimulus levels will

have a finite mean. Therefore, Tsutakawa's formulas can be applied. Table I lists the asymptotic distributions that were calculated for different combinations of k, s, and r, along with their means and variances. Since $\hat{\mu}$ converges in probability to the asymptotic mean, for $\mu = 0$ this mean is the asymptotic bias of the block up-and-down procedure (as $n \to \infty$).

III. RESULTS AND DISCUSSION

Formulas 1 and 2 were used to compute exact bias and mean-square error (MSE) of the threshold estimator $\hat{\mu}$. The parameters that determine bias and precision of $\hat{\mu}$ are: (1) the block design (k,s,r), (2) the initial stimulus level y_o, (3) the step size d, (4) the number of trial blocks n, and (5) the phasing factor. The phasing factor was discussed earlier and eliminated from consideration. Although the remaining four parameters are under the control of the experimenter, he seldom has enough prior information about the true values of μ and σ to allow him to select y_o, d, and n optimally. This study concentrates, therefore, on finding block designs that are good over wide ranges of y_o, d, and n. Some sample curves of bias and MSE are included, in special cases, to provide some insight into the relative influence of these parameters.

The block design (k,s,r) will be considered first. Not all combinations (k,s,r) are feasible. Since the up-and-down method concentrates testing around a so-called target percentage correct, we must see what percentile of the function P(x) is actually tracked by the

procedure. If the two tail probabilities of the binomial distribution discussed in the previous section are equated, and the resulting equation is solved for the binomial parameter p, the result is the probability of correct response tracked by the up-and-down procedure. Table II gives this value of p as a function of (k,s,r) along with the normalized stimulus level that yields the probability p; that is, the value of z is found by numerically solving the equation

$$\frac{1}{2}\left[1 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt\right] = p \qquad (8)$$

for z, given p. The stimulus level that yields p is then $x = \mu + z\sigma$. Because of the asymmetry of the response function, x is not exactly equal to the asymptotic value of $\hat{\mu}$ listed in Table I. However, these two numbers are expected to be close for suitable step sizes d; combinations (k,s,r) chosen from Table II for further investigation were those that yielded values close to 0.75 for p. As shown in the derivation of $P(x)$, the inequality of $0.5 < p < 1$ must hold, which accounts for the missing entries (k,s,r) in the table. The futility of considering combinations (k,s,r) that do not appear in Table II can also be shown by the asymptotic behavior of the up-and-down series in this case. It can easily be shown that for $p < 0.5$, where p is the probability that equates the two tail probabilities discussed above, the probability of the stimulus level x decreasing on the next trial following a given trial is greater than the probability of increasing (i.e., $P_- > P_+$). Therefore, the

process tends to drift to $-\infty$, and the absolute bias and MSE increase without bound.

A reasonable way to choose the best s and r for a given block size, k, or to choose the best k for a fixed total number of trials, is to base comparisons on the amount of information (reciprocal of MSE) per trial.[13] This measure is called the efficiency E:

$$E = \frac{1}{nk(MSE)} \tag{9}$$

where k is the block size, n is the number of trial blocks, and MSE is the observed mean-square error averaged over all initial stimulus levels, y_0. Figures 1 through 3 give E as a function of the total trial sequence length nk for step sizes d = 1/2, 1, and 2. The values of s and r chosen for each k were those that gave the highest efficiencies over the range of trial lengths considered. Whenever no single combination of s and r was uniformly best for all trial lengths, one of the two best was plotted. The figures show very strikingly the relatively poor performance of k = 2, the value of k very often used in psychometric work, especially for long sequence lengths and large step size. On the other hand, the uniformly good performance of k = 5 is surprising. Note that a block size of 5 would be ruled out if the criterion stated on page 1177 of Reference 2 were followed (the rule that it must be possible for the proportion of correct responses out of the block of five to be exactly equal to 0.75).

Another way of looking at the up-and-down method is to compare its performance to that of a fixed stimulus level design (e.g., the

classical z-score method of threshold estimation). A fixed-level method very similar to the z-score method, and that has served as a basis for comparison with the up-and-down method in bioassay,[10,12,13] is the method of probits developed by Finney.[16] This method is based on a maximum likelihood estimate of threshold, and is therefore known to be asymptotically efficient. The fixed-level design chosen was based on an equal number of trials at each of five stimulus levels, and the formula for the variance of $\hat{\mu}$ given in Reference 16 was used, with σ assumed to be 1 and the weights adjusted to conform to the response function in this report, P(x). The relative efficiency (RE) of the BUDTIF method is defined as

$$\frac{\text{MSE (PROBIT)}}{\text{MSE (BUDTIF)}} \qquad (10)$$

Representative efficiency curves in Fig. 4 show that the BUDTIF method is relatively efficient for starting levels away from threshold and for small trial sequence lengths. Even for starting levels at threshold, the BUDTIF method seems to be more efficient than the probit estimator for large block sizes. Relative efficiencies were averaged over starting levels, y_o, for various (k,s,r) and are shown in Table III for step size d = 1.

Bias and MSE of the estimator $\hat{\mu}$ are shown in Fig. 5 for k = 8 and d = 1.5. Mean-square error based on a constant total trial sequence length (nk = 20) is plotted in Fig. 6 for various combinations (k,s,r). It is obvious that starting stimulus levels far away

from threshold inflate the bias and MSE; however, it is not uniformly true that threshold is the best starting position. Because of the particular response function assumed, the bias for starting levels at threshold tends to be negative and becomes worse as the trial sequence length increases. The block size of $k = 8$, however (Fig. 5), is remarkably stable and has a uniformly small positive bias up to moderate sequence lengths. Table IV shows that the asymptotic mean stimulus level is negative for most values of (k,s,r), explaining why the bias goes negative with increasing trial sequence lengths.

The effect of step size d, as shown in Fig. 6, is similar to that previously shown for the bioassay case. If y_o is far from threshold, the MSE increases with d. A small d gives a very precise threshold estimate if y_o is close to μ, but a large d is more stable with respect to MSE as y_o gets farther from μ.

IV. CONCLUDING REMARKS

Much of the past research on sampling properties of statistical estimates of sensitivity thresholds has been based on lengthy and inaccurate Monte Carlo computations that required repeated sampling of the responses of a simulated human subject. Although the Monte Carlo approach is necessary for analyzing some of the proposed sequential procedures that are analytically intractable, the methods described in this paper are adequate for computing exact bias and precision for fixed-length block designs in which an average estimator is used. This makes possible the rapid search of many more parameters than is practical with Monte Carlo procedures.

This investigation of the block-up-and-down design has shown that the design often used in the past, based on a trial sequence length of two, is one of the worst from the standpoint of sampling efficiency. Efficiency curves were provided to assist the experimenter in choosing an appropriate block design.

REFERENCES

[1]E. G. Heinemann, "The Relation of Apparent Brightness to the Threshold for Differences in Luminance," J. Exp. Psychol. 61(5), 389-399 (1961).

[2]C. W. Stuckey, C. L. Hutton, and R. Campbell, "Decision Rules in Threshold Determination," J. Acoust. Soc. Amer. 40(5), 1174-1179 (1966).

[3]T. N. Cornsweet, "The Staircase Method in Psychophysics," Amer. J. Psychol. 75, 485-491 (1962).

[4]S. Feeny, P. K. Kaiser, and J. P. Thomas, "An Analysis of Data Gathered by the Staircase Method," Amer. J. Psychol. 79, 652-654 (1966).

[5]B. Clark and J. D. Stewart, "Comparison of Three Methods to Determine Thresholds for Perception of Angular Acceleration," Amer. J. Psychol. 81, 207-216 (1968).

[6]M. M. Taylor and C. D. Creelman, "PEST: Efficient Estimates on Probability Functions," J. Acoust. Soc. Amer. 41(4), 782-787 (1967).

[7]R. A. Campbell and E. Z. Lasky, "Adaptive Threshold Procedures: BUDTIF," J. Acoust. Soc. Amer. 44(2), 537-541 (1968).

[8]R. A. Campbell, "Detection of a Noise Signal of Varying Duration," J. Acoust. Soc. Amer. 35(11), 1732-1737 (1963).

[9]G. B. Weatherill, "Sequential Estimation of Quantal Response Curves," J. Roy. Statist. Soc., Ser. B 25, 1-48 (1961).

[10]W. J. Dixon and A. M. Mood, "A Method for Obtaining and Analyzing Sensitivity Data," J. Amer. Statist. Assoc. 43, 109-126 (1948).

[11]R. K. Tsutakawa, "The Random Walk Design in Bio-Assay," J. Amer. Statist. Assoc. 62(319), 842-856 (1967).

[12]B. P. Hsi, "The Multiple Sample Up-and-Down Method in Bioassay," J. Amer. Statist. Assoc. 64(325), 147-162 (1969).

[13]K. A. Brownlee, J. L. Hodges, and M. Rosenblatt, "The Up-and-Down Method With Small Samples," J. Amer. Statist. Assoc. 48, 262-277 (1953).

[14]G. B. Wetherill and H. Levitt, "Sequential Estimation of Points on a Psychometric Function," British J. Math. Statist. Psychol. 18, 1-10 (1966).

[15]M. M. Taylor and C. D. Creelman, "PEST: A Rapid Technique for Finding Arbitrary Points on a Psychometric Function," Psychonom. Soc., Chicago (1965).

[16]D. J. Finney, *Probit Analysis* (Cambridge Univ. Press, 1952).

TABLE I. Asymptotic frequency distributions of stimulus levels for step size d = 1. Triplets in parentheses are (k,s,r); phasing factor ϕ = 0.0.

Level	(2,1,2)	(3,2,3)	(4,2,4)	(5,3,5)	(6,3,6)	(7,5,6)	(8,4,8)	(9,4,9)	(10,6,9)
4	0.000001	0.000004	0.000000	0.000000	0.000000	0.000000			
3	.000601	.002209	.000005	.000020	.000000	.000072	0.000000	0.000000	0.000000
2	.026542	.065331	.006311	.016087	.001493	.027539	.000357	.000026	.000939
1	.170262	.287458	.177920	.283830	.167637	.271655	.152473	.076366	.165937
0	.329875	.388033	.488438	.511324	.602561	.439936	.691567	.741818	.603113
-1	.279290	.203213	.275881	.173884	.215138	.225835	.152419	.178990	.220279
-2	.126927	.045611	.046365	.014220	.012815	.032362	.003163	.002788	.009581
-3	.044296	.006975	.004619	.000613	.000350	.002428	.000023	.000013	.000150
-4	.014806	.001000	.000422	.000024	.000008	.000163	.000000	.000000	.000002
-5	.004935	.000143	.000038	.000001	.000000	.000011			.000000
-6	.001645	.000020	.000003	.000000		.000001			
-7	.000548	.000003	.000000			.000000			
-8	.000183	.000000							
-9	.000061								
-10	.000020								
-11	.000007								
-12	.000002								
Var.	1.6412	1.0267	0.6764	0.5725	0.4382	0.7617	0.3192	0.2550	0.4245
Mean	-0.5408	0.1045	-0.1938	0.1118	-0.07123	0.02839	-0.005628	-0.1082	-0.07208

TABLE II. True probability, p, of correct response tracked by the design (k,s,r); stimulus level that yields this p is z.

k	s	r	p	z	k	s	r	p	z
2	1	2	0.7071	-0.2167		6	7	0.7989	0.2475
3	1	3	.6527	-.5089			8	.8620	.5947
	2	3	.7937	.2209		7	8	.9170	.9701
4	1	4	.6245	-.6775	9	1	9	.5707	-1.0742
	2	3	.6143	-.7437		2	8	.5600	-1.1748
		4	.7336	-.0822			9	.6301	-.6427
	3	4	.8409	.4727		3	7	.5555	-1.2211
5	1	5	.6066	-.7957			8	.6150	-.7386
	2	4	.5943	-.8827			9	.6839	-.3375
		5	.6980	-.2636		4	6	.5537	-1.2408
	3	4	.6862	-.3256			7	.6087	-.7808
		5	.7839	.1707			8	.6673	-.4274
	4	5	.8706	.6467			9	.7344	-.0784
6	1	6	.5939	-.8864		5	6	.6069	-.7932
	2	5	.5815	-.9818			7	.6611	-.4618
		6	.6736	-.3928			8	.7181	-.1607
	3	4	.5786	-1.0061			9	.7827	.1646
		5	.6587	-.4752		6	7	.7138	-.1827
		6	.7472	-.0143			8	.7686	.0934
	4	5	.7355	-.0725			9	.8298	.4115
		6	.8182	.3488		7	8	.8204	.3605
	5	6	.8909	.7783			9	.8768	.6860
7	1	7	.5843	-.9597		8	9	.9259	1.0440
	2	6	.5724	-1.0588	10	1	10	.5656	-1.1206
		7	.6555	-.4929		2	9	.5556	-1.2207
	3	5	.5684	-1.0948			10	.6207	-.7016
		6	.6398	-.5837		3	8	.5511	-1.2694
		7	.7206	-.1479			9	.6062	-.7978
	4	5	.6359	-.6075			10	.6705	-.4096
		6	.7053	-.2258		4	7	.5489	-1.2937
		7	.7824	.1633			8	.5997	-.8439
	5	6	.7715	.1080			9	.6541	-.5009
		7	.8431	.4850			10	.7171	-.1659
	6	7	.9057	.8832		5	6	.5483	-1.3011
8	1	8	.5768	-1.0213			7	.5969	-.8638
	2	7	.5655	-1.1217			8	.6470	-.5418
		8	.6414	-.5743			9	.7003	-.2518
	3	6	.5611	-1.1642			10	.7614	.0573
		7	.6259	-.6688		6	7	.6449	-.5540
		8	.7002	-.2522			8	.6939	-.2850
	4	5	.5598	-1.1765			9	.7457	-.0218
		6	.6203	-.7043			10	.8043	.2759
		7	.6838	-.3384		7	8	.7414	-.0430
		8	.7556	.0280			9	.7910	.2072
	5	6	.6795	-.3612			10	.8465	.5044
		7	.7408	-.0462		8	9	.8377	.4551
		8	.8090	.3004			10	.8888	.7640
						9	10	.9330	1.1080

TABLE III. Relative efficiency of the average estimator with respect to the probit estimator averaged over all starting levels y_o; d = 1; RE = MSE(PROBIT)/MSE(BUDTIF).

	n				
(k,s,r)	5	10	20	25	50
(2,1,2)	4.36	3.03	1.72	1.47	0.94
(4,2,4)	3.76	3.63	---	3.22	---
(5,3,5)	3.56	4.10	5.03	---	---
(6,4,6)	2.93	---	---	---	---
(8,4,8)	2.19	---	---	---	---
(10,6,9)	---	3.26	---	---	---

TABLE IV. Means and variances of the asymptotic frequency distribution of stimulus levels.

	d = 1/2		d = 1		d = 1-1/2	
(k,s,r)	Mean	Var.	Mean	Var.	Mean	Var.
(2,1,2)	-0.3552	0.6142	-0.5408	1.6412	-0.7557	3.1476
(3,2,3)	.1686	.4210	.1045	1.0267	.0320	1.838
(4,2,4)	-.1306	.2737	-.1938	.6764	-.2653	1.254
(5,3,5)	.1448	.2340	.1118	.5725	.0853	1.056
(6,3,6)	-.03663	.1784	-.07123	.4382	-.1132	.8072
(7,5,6)	.0691	.3081	.02839	.7617	.0014	1.399
(8,4,8)	.0188	.1327	-.005628	.3192	-.0368	.5658
(9,4,9)	-.0862	.1127	-.1082	.2550	-.1390	.4335
(10,6,9)	-.0446	.1649	-.07208	.4245	-.0943	.8293

FIGURE LEGENDS

Fig. 1. Efficiency $E = 1/kn\ (MSE)$ versus trial sequence length (kn) for step size $d = 0.5$. Mean-square error (MSE) was averaged over all starting stimulus levels y_o. Numbers in parentheses are (k,s,r).

Fig. 2. Efficiency $E = 1/kn\ (MSE)$ versus trial sequence length (kn) for step size $d = 1$. Mean-square error (MSE) was averaged over all starting stimulus levels y_o. Numbers in parentheses are (k,s,r).

Fig. 3. Efficiency $E = 1/kn\ (MSE)$ versus trial sequence length (kn) for step size $d = 2$. Mean-square error (MSE) was averaged over all starting stimulus levels y_o. Numbers in parentheses are (k,s,r).

Fig. 4. Relative efficiency (RE) of the average estimator with respect to the probit estimator for $nk = 40$. Triplets in parentheses are (k,s,r) and $d = 1$.

Fig. 5. (a) Bias of the estimator $\hat{\mu}$ versus initial stimulus y_o and sample size n; $(k,s,r) = (8,4,8)$ and $d = 1.5$. (b) MSE of the estimator $\hat{\mu}$ versus initial stimulus y_o and sample size n; $(k,s,r) = (8,4,8)$ and $d = 1.5$.

Fig. 6. Mean-square error of the average estimator $\hat{\mu}$ for constant trial sequence length ($nk = 20$) for $k = 2$, 4, and 10. Numbers in parentheses denote (k,s,r).

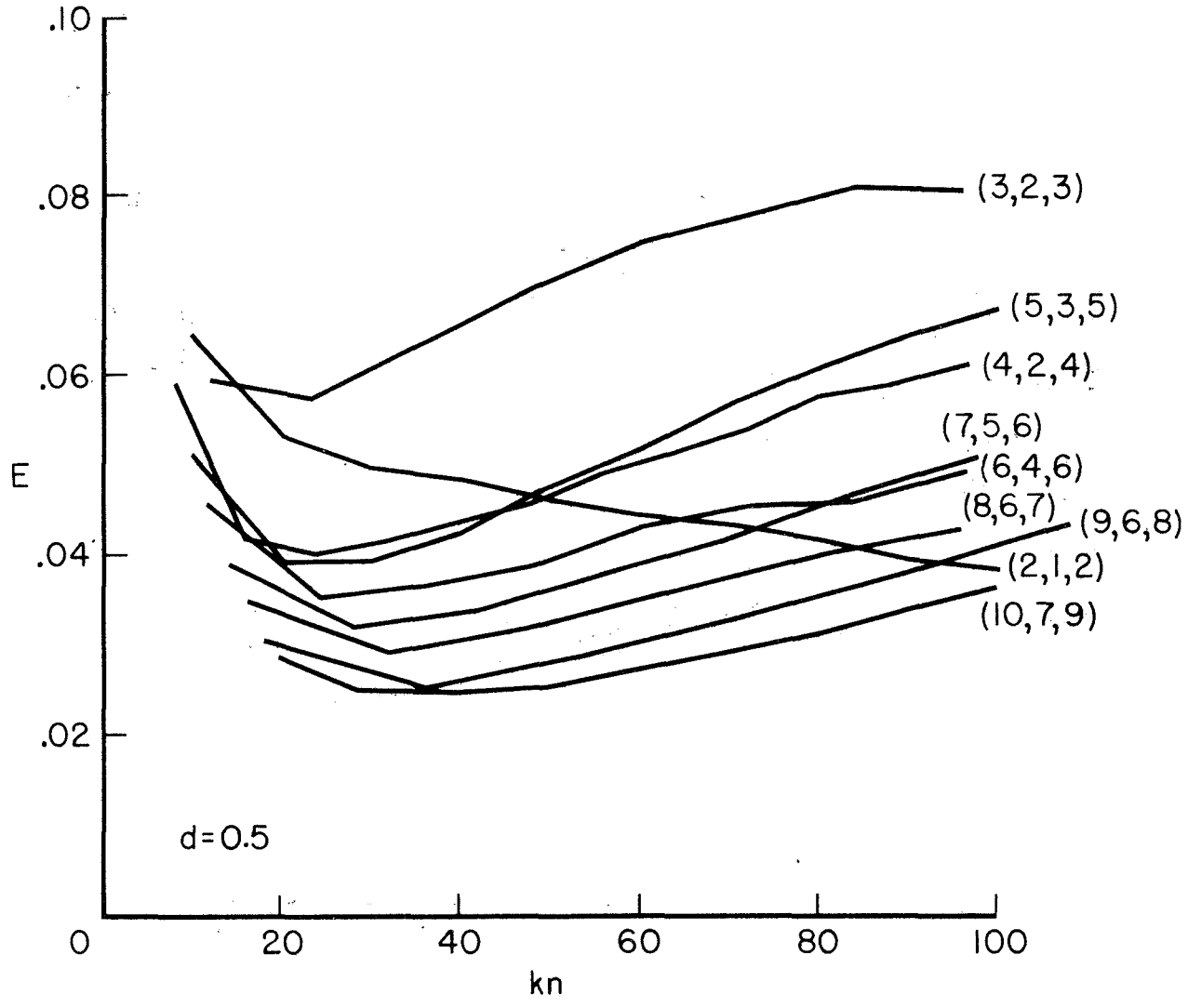
.10
.08
.06
.04
.02
0
E
20
40
60
80
100
kn
d=0.5
(3,2,3)
(5,3,5)
(4,2,4)
(7,5,6)
(6,4,6)
(8,6,7)
(9,6,8)
(2,1,2)
(10,7,9)

Figure 1

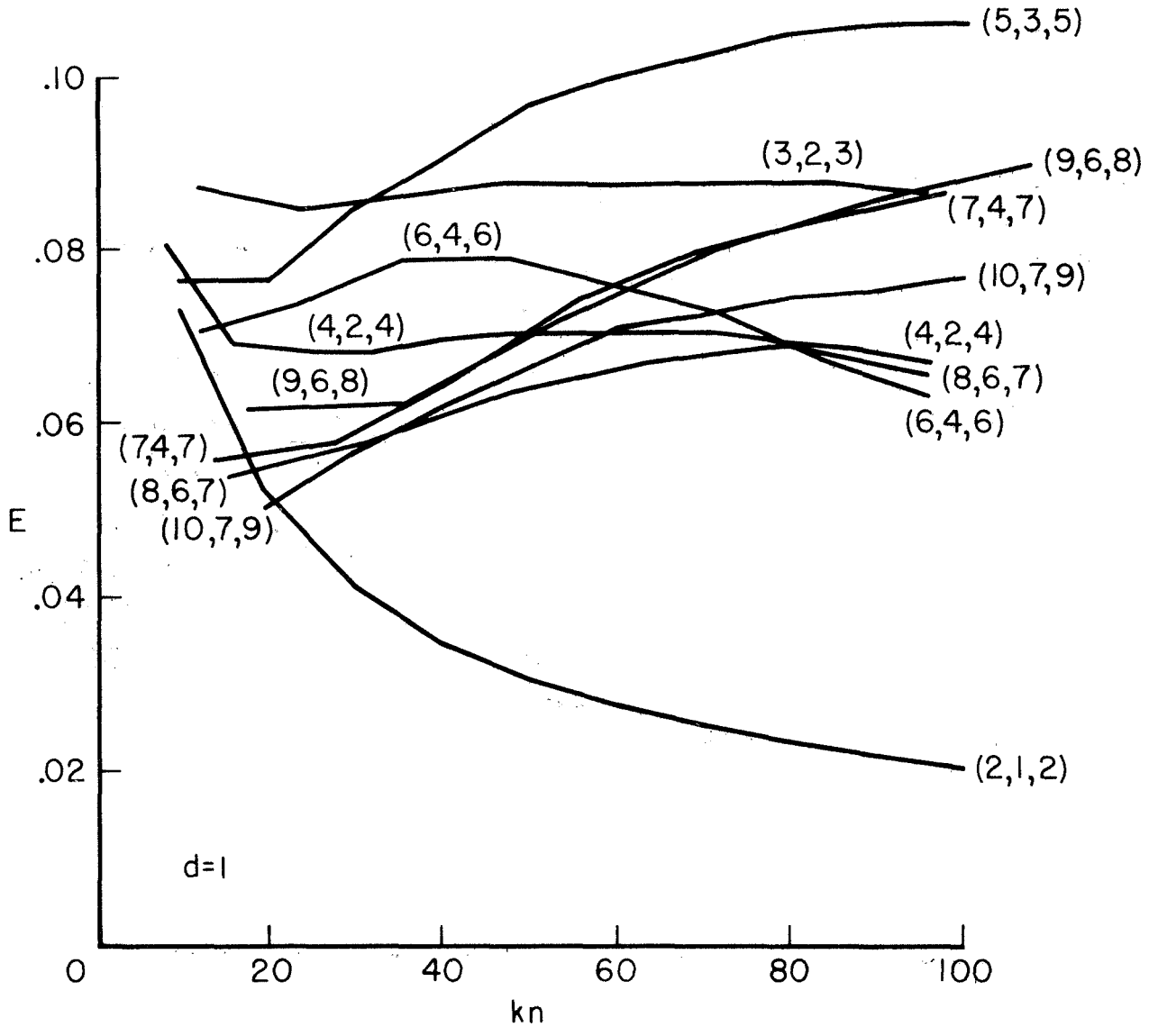
(5,3,5)
(3,2,3)
(9,6,8)
(7,4,7)
(6,4,6)
(10,7,9)
(4,2,4)
(4,2,4)
(9,6,8)
(8,6,7)
(6,4,6)
(7,4,7)
(8,6,7)
(10,7,9)
(2,1,2)
.10
.08
.06
.04
.02
E
d=1
0
20
40
60
80
100
kn

Figure 2

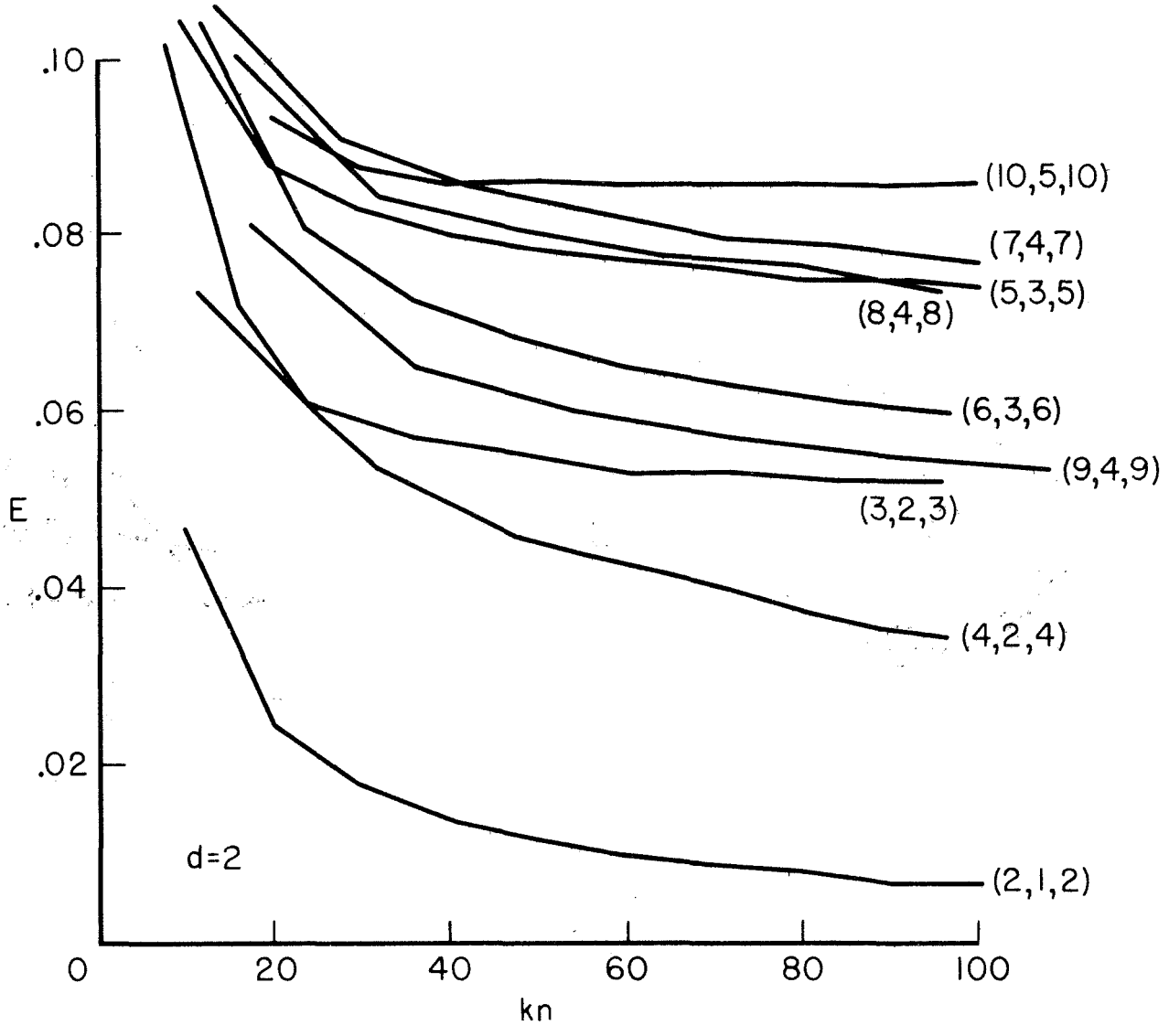
.10
.08
.06
.04
.02
E
0
20
40
60
80
100
kn
d=2
(10,5,10)
(7,4,7)
(5,3,5)
(8,4,8)
(6,3,6)
(9,4,9)
(3,2,3)
(4,2,4)
(2,1,2)

Figure 3

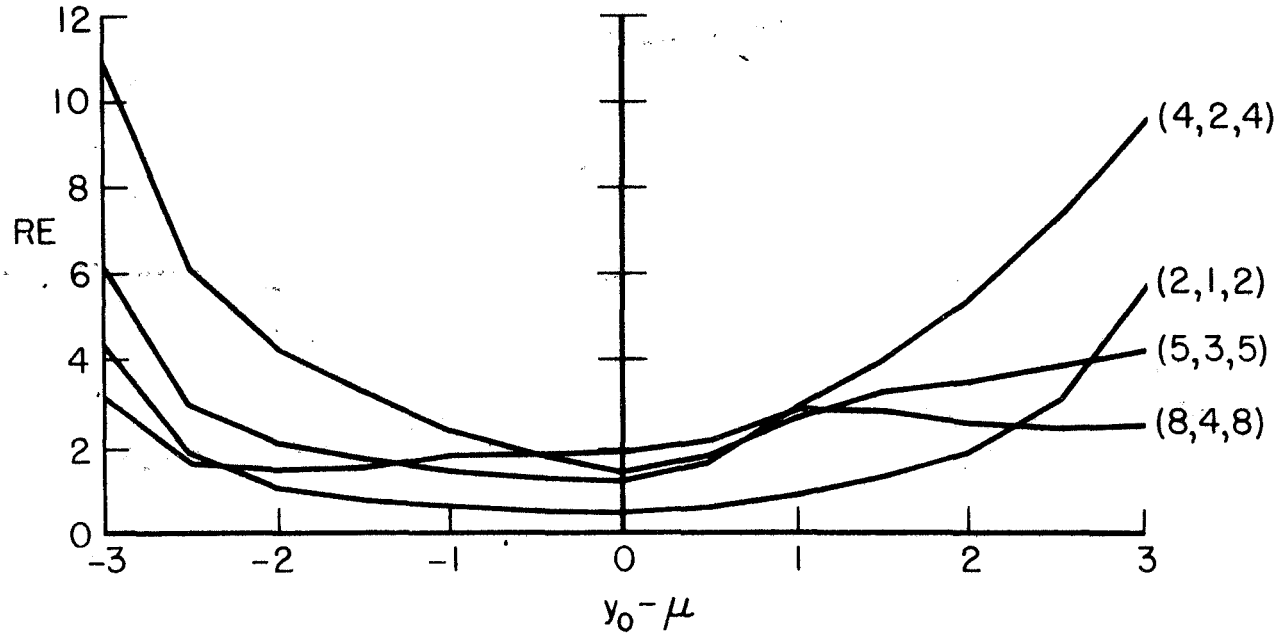

Figure 4

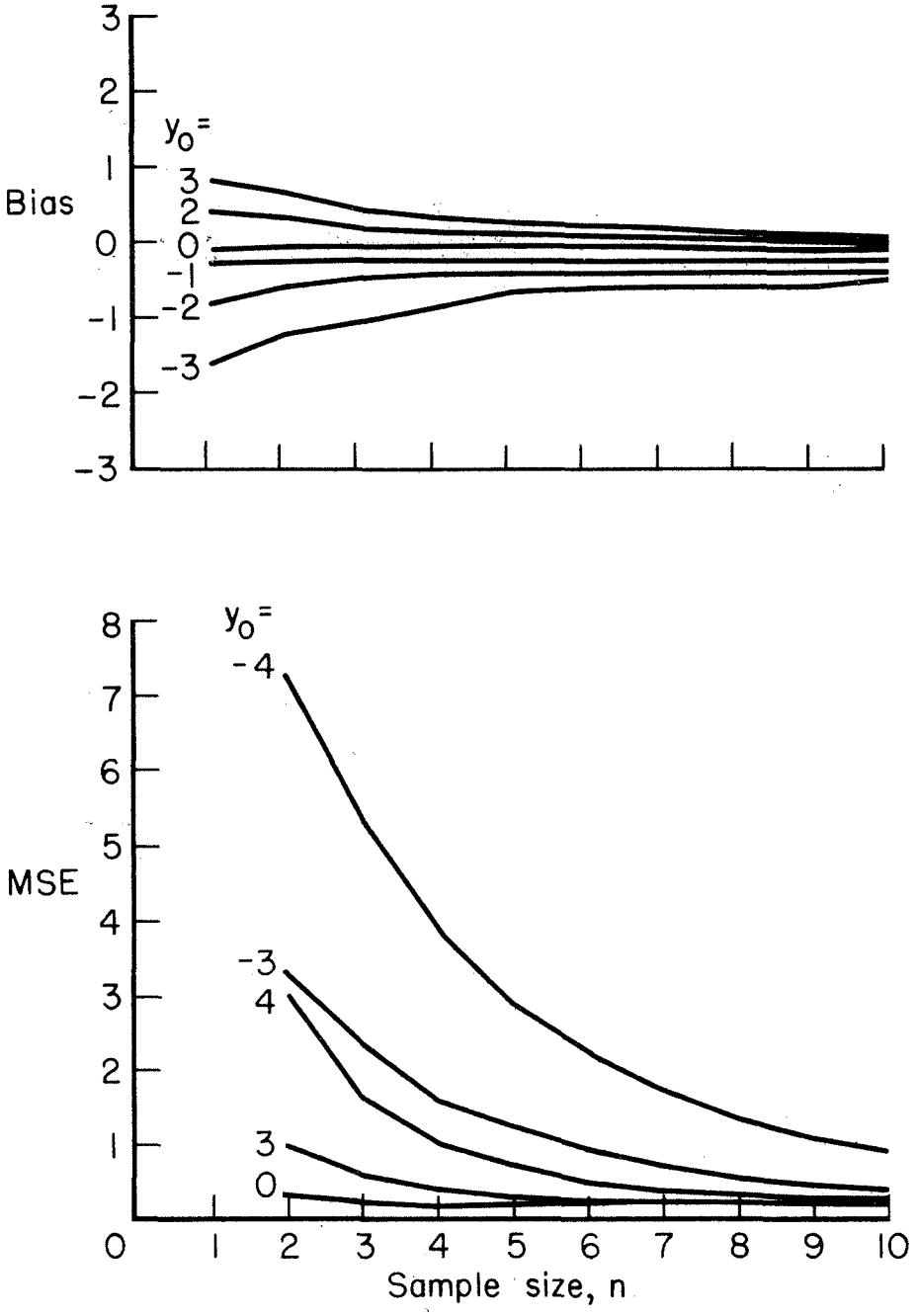

Figure 5

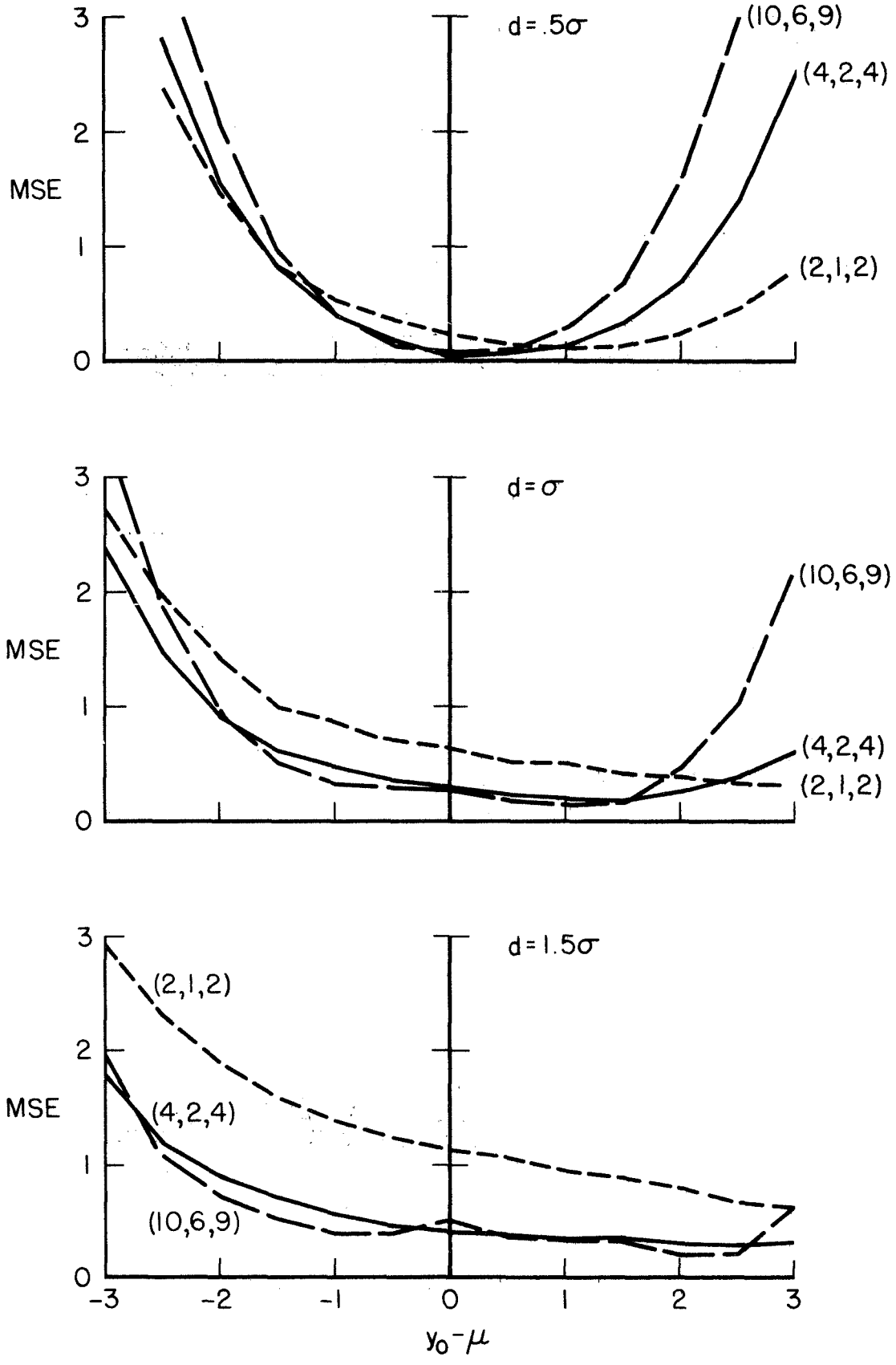
d = .5σ
(10,6,9)
(4,2,4)
(2,1,2)
MSE
d = σ
(10,6,9)
(4,2,4)
(2,1,2)
MSE
d = 1.5σ
(2,1,2)
(4,2,4)
(10,6,9)
MSE
0
1
2
3
-3
-2
-1
0
1
2
3
$y_0 - \mu$

Figure 6